

Serial No.: 10/763,614
Attorney Docket: 3394.2

AMENDMENTS

Amendments to the Specification

Please replace the paragraph on page 1 under the section heading "Cross Reference to Related Applications" with the following amended paragraph:

The present application is a continuation of U.S. Patent Application Serial No. 09/683, 710, filed on February 5, 2002, which inventors claim claims priority to U.S. Provisional Application Serial No. 60/266,718, filed February 2, 2001, which is on February 5, 2001. Both applications are hereby incorporated by reference in its their entirety for all purposes.

Please replace paragraph [0003] on page 2 with the following amended paragraph:

In one aspect of the invention, methods are provided for detecting a transcribed genomic region. The methods include providing a nucleic acid sample containing transcripts or nucleic acids ~~derried~~ derived from transcripts from the genome; hybridizing the nucleic acid sample with a plurality of nucleic acid probes, where the probes are designed to ~~interogate~~ interrogate potential transcripts from both strands of the genomic DNA; and analyzing hybridization signals to detect the transcribed region.

Please replace paragraph [0004] on page 2 with the following amended paragraph:

In some ~~embodiments~~, the ~~pluarlity~~ plurality of probes comprises probes ~~interogating~~ interrogating the intergenic, and intronic regions of the genome. The probes may be immobilized on a substrate at a density greater than 400 or 1000 different probes per cm².

Please replace paragraph [0008] on page 2 with the following amended paragraph:

In yet another aspect of the invention, methods for detecting an untranslated region (UTR) for a gene are provided. The methods include hybridizing a sample containing transcripts or nucleic acids ~~derried~~ derived from transcripts with a plurality of

Serial No.: 10/763,614
Attorney Docket: 3394.2

probes, where the probes interrogate transcription of an intergenic region immediately upstream the gene; and classifying the intergenic region as a potential ~~5"UTR~~ 5'UTR of the gene if the intergenic region is transcribed in the same orientation of the gene and the ~~transcribed~~ transcribed region is greater than 70 bases in length. Similarly, an intergenic region is classified as a potential ~~3"UTR~~ 3'UTR of the gene if the intergenic region is transcribed in the same orientation of the gene, it is immediately downstream of the gene and the ~~transcribed~~ transcribed region is greater than 70 bases in length.

Please replace paragraph [0011] on page 3 with the following amended paragraph:

Figure 2 shows ~~5"UTR~~ 5'UTR detection upstream of *ompA*. Individual oligonucleotide probe intensities (PM MM) from three conditions to validate the microarray detected ~~5"UTR~~ 5'UTR upstream of *ompA* (22). Intensities for individual oligonucleotide probes interrogating *ompA*, the 356 bp *Ig* region and *galU* are shown. The arrows above the indicated genes show the direction of transcription.

Please replace paragraph [0029] on page 8 with the following amended paragraph:

One of skill in the art would appreciate that it is desirable to inhibit or destroy ~~RNase~~ RNases present in homogenates before homogenates can be used for hybridization. Methods of inhibiting or destroying nucleases are well known in the art. In some preferred embodiments, cells or tissues are homogenized in the presence of chaotropic agents to inhibit nuclease. In some other embodiments, ~~RNase~~ RNases are inhibited or destroyed by ~~heat~~ heat treatment followed by proteinase treatment.

Please replace paragraph [0031] on page 9 with the following amended paragraph:

In a preferred embodiment, the total RNA is isolated from a given sample using, for example, an acid guanidinium-phenol-chloroform extraction method and polyA⁺ mRNA is isolated by oligo dT column chromatography or by using (dT)_n magnetic beads. (See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989), or *Current Protocols in Molecular*

Serial No.: 10/763,614
Attorney Docket: 3394.2

Biology, F. Ausubel et al., ed. Greene Publishing and Wiley-Interscience, New York, 1987.) In one particularly preferred embodiment, total RNA is isolated from mammalian cells using RNeasy Total RNA isolation kit (QIAGEN). If mammalian tissue is used as the source of RNA, a commercial reagent such as TRIzol Reagent (GIBCOL Life Technologies) may be used. A second cleanup after the ethanol precipitation step in the TRIzol extraction using Rneasy total RNA isolation kit may be beneficial.

Please replace paragraph [0034] on page 9 with the following amended paragraph:

Total RNA from prokaryotes, such as ~~E. coli~~ E. coli cells, may be obtained by following the protocol for MasterPure complete DNA/RNA purification kit from Epicentre Technologies (Madison, WI).

Please replace paragraph [0037] on page 10 with the following amended paragraph:

In a particularly preferred embodiment, the sample mRNA is reverse transcribed with a reverse transcriptase and a primer consisting of oligo dT and a sequence encoding the phage T7 promoter to provide a single stranded DNA template. The second DNA strand is polymerized using a DNA polymerase with or without primers. (See U.S. Patent Application Serial Number: 09/102,167, and U.S. ~~Provisional Application Serial No. 60/172,340~~ Patent Application Serial Number 10/763,414, both incorporated herein by reference for all purposes.) After synthesis of double-stranded cDNA, T7 RNA polymerase is added and RNA is transcribed from the cDNA template. Successive rounds of transcription from each single cDNA template results in amplified RNA. Methods of in vitro polymerization are well known to those of skill in the art. (See, e.g., Sambrook, supra.) and this particular method is described in detail by Van Gelder et al., Proc. Natl. Acad. Sci. USA, 87: 1663-1667, 1990. Moreover, Eberwine et al. Proc. Natl. Acad. Sci. USA, 89: 3010-3014 provide a protocol that uses two rounds of amplification via in vitro transcription to achieve greater than 10⁶ fold amplification of the original starting material thereby permitting expression monitoring even where biological samples are limited. In one preferred embodiment, the in-vitro transcription reaction may be coupled with labeling of the resulting cRNA with biotin using Bioarray high yield RNA transcript labeling kit (Enzo P/N 900182).

Serial No.: 10/763,614
Attorney Docket: 3394.2

Please replace paragraph [0046] on page 14 with the following amended paragraph:

In some embodiments, nucleic acid probes designed to detect transcripts from a region of a genome are hybridized with a nucleic acid sample derived from the species with the genome. Because either strand of the genomic DNA can serve as a template, probes that can detect the transcripts or nucleic acids ~~derived~~ derived from the transcripts may be employed. Methods for deciphering which strand ~~act~~ acts as the template for a transcript are described in, for example, U.S. Patent Application Serial Number 09/683,221, filed on 12/3/2001, which issued as U.S. Patent No. 6,670,122 which is incorporated herein by reference for all purposes. In some embodiments, the actual sequences of the nucleic acid probes may be dependent upon the assay protocols. For example, if the transcripts are directly hybridized, the probes for detecting the transcripts should be complementary potential transcripts. Alternatively, if a sample is derived from the transcripts, via, for example, reverse transcription or amplification, the probes should be complementary with the derived nucleic acids. The probes may be designed according to the reference sequence of a genome. In a particularly preferred embodiment, probe sequences are obtained from both ~~strand~~ strands of the genomic DNA so that potential transcripts from either strand can be detected.

Please replace paragraph [0047] on page 15 with the following amended paragraph:

While various aspects of the invention are primarily described using ~~exemplary~~ exemplary embodiments which use high density oligonucleotide probes, this invention is not limited to any particular microarray format. For example, the probes may be presynthesized, and immobilized on beads or optical fibers.

Please replace paragraph [0062] on page 20 with the following amended paragraph:

In another aspect of the invention, methods are provided for detecting an operon element in a prokaryote. The methods include hybridizing transcripts or nucleic acids ~~derived~~ derived from transcripts from the organism with a plurality of probes, where the probes interrogate transcription of an intergenic region between two flanking open

Serial No.: 10/763,614
Attorney Docket: 3394.2

reading frames (ORFs); and classifying the intergenic region as a potential operon element if both flanking ORFs are expressed and if the intergenic region is transcribed off the same DNA strand as the flanking ORFs.

Please replace paragraph [0064] on page 20 with the following amended paragraph:

In some preferred embodiments, ~~method~~ the methods include classifying the intergenic region as a potential operon element if both flanking ORFs are expressed and if the intergenic region is transcribed off the same DNA strand as the flanking ORFs and the transcription of the intergenic region is correlated with the transcription of at least one of the flanking ORFs.

Please replace paragraph [0065] on page 20 with the following amended paragraph:

In yet another aspect of the invention, methods for detecting untranslated region (UTR) for a gene are provided. The methods include hybridizing a sample containing transcripts or nucleic acids ~~derived~~ derived from transcripts with a plurality of probes, where the probes interrogate transcription of an intergenic region immediately upstream the gene; and classifying the intergenic region as a potential ~~5'UTR~~ 5'UTR of the gene if the intergenic region is transcribed in the same orientation ~~of the~~ as the gene and the ~~transcribed~~ transcribed region is greater than 70 bases in length. Similarly, an intergenic region is classified as a potential ~~3'UTR~~ 3'UTR of the gene if the intergenic region is transcribed in the same orientation ~~of the~~ as the gene, ~~it is~~ is immediately downstream of the gene and the ~~transcribed~~ transcribed region is greater than 70 bases in length.

Please replace paragraph [0066] on page 21 with the following amended paragraph:

This example (See, Brian Tjaden, 2001, Transcriptome Analysis of *Escherichia coli* using High-Density Oligonucleotide Probe Arrays, *Genes & Development*, 15:1637, incorporated herein by reference for all purposes) shows the interrogation of the *Escherichia coli* MG1655 genome sequence for transcription activities and the identification of transcripts according to the exemplary embodiments of the invention.

Serial No.: 10/763,614
Attorney Docket: 3394.2

By interrogating both strands of a genome sequence on a microarray at a high resolution, RNA transcripts can be globally identified and linked back to the genome sequence, allowing more accurate annotation predictions. In this study, high-density oligonucleotide probe arrays on which the complete *Escherichia coli* MG1655 genome sequence is represented were used to identify RNA transcripts in the intergenic (Ig) regions. Each previously annotated open-reading frame (ORF) (Blatner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12 [see comments]. *Science* 277, 1453-74 (1997)) has 15 oligonucleotide probes, which are designed to be complementary to the sense strand and each intergenic region greater than 40 bp is interrogated with 15 probes on each of the forward and reverse strands. Since microarrays traditionally interrogate only the *in silico* identified translated region of a gene, we consider all elements between translated regions as intergenic. The sequence of the oligonucleotide probes and their location in regards to the genome sequence have been published (arop.med.harvard.edu/ExpressDB/EDS37/GAPS_webpages/GAPS_main.htm, last visited on Feb. 2, 2002) on the website of the Harvard-Lipper Center for Computational Genetics and provide the basis for a detailed analysis of the *E. coli* transcriptome.

Please replace paragraph [0073] on page 24 with the following amended paragraph:

Ig transcripts are classified as operon elements if both flanking ORFs are expressed, if the Ig region is transcribed off the same DNA strand as the flanking ORFs and if the expressed transcript extends across the entire Ig region, except possibly isolated single probes. To improve sensitivity, we allow up to one probe in a probe set not to be expressed. Using these criteria, 293 transcripts and their flanking genes were identified as operon elements. 289 of these Ig regions have been previously documented or predicted as being part of an operon (http://www.cifn.unam.mx/Computational_Genomics/GETools/E.coli_predictions.html) (see for example, Gene Expression Analysis tools or GETtools on the website of the Nitrogen Fixation Centre of the National Autonomous University of Mexico). Based on this comparison the false positive rate for transcript detection was estimated to be less than 1%. Cluster analysis revealed that 71% of the previously predicted operons showed

Serial No.: 10/763,614
Attorney Docket: 3394.2

co-regulation of at least two out of three transcripts (flanking genes and Ig region) while 81% of the documented operons offered this evidence of co-regulation. When co-regulation for all three transcripts was required, 17% of the predicted operons showed evidence compared to 44% of the documented operons. Figure 1 shows the expression levels for individual probes interrogating the predicted *hnr-galU* operon. RT-PCR confirmed a single RNA transcript for these two genes and the Ig region (data not shown). Six additional operons were experimentally confirmed using RT-PCR (Table 3, supplemental data). From a total of 931 predicted and documented operons in Regulon DB (24) which meet our criteria for being operon elements, we detect 334 using our microarray analysis. This results in a false negative rate of less than 64%. This unusual high false negative rate is consistent with the fact that we use a very conservative transcript prediction model and in addition the majority of the operons listed in Regulon DB are predicted operons without experimental validation. Two Ig regions that have not been reported to be part of an operon were found to be co-regulated either with both flanking genes (C0794: *rpsM/rpmJ*) or with the downstream gene (C0789: *rplN/rpsQ*). Both Ig regions are flanked on one side by documented operons containing genes for 30S and 50S ribosomal subunit proteins and on the other side with a gene encoding a 50S ribosomal subunit protein. Based on our findings and the close functional relationship of the gene products, they are strong candidates for new, previously unidentified operons. The third potential operon candidate (C0669; *nlpD/pcm*) was found to have co-regulated flanking genes. The two genes have no obvious functional relationships and need to be further analyzed. The fourth operon candidate (C0064: *yaeD/rrsH*) shows no co-regulation with the flanking genes and is located between a gene with unknown function and the 16S RNA of the *rrnH* operon.

Please replace paragraph [0074] on page 25 with the following amended paragraph:

As with the operons described above, experimental evidence for 5-prime expressed regions can supplement computational approaches by identifying not only transcription start sites for genes, but also multiple start sites when different promoters are employed under different conditions as well as *cis*-regulatory sites upstream of known

Serial No.: 10/763,614
 Attorney Docket: 3394.2

genes. In order for an Ig transcript to be classified as a 5' UTR 5' UTR in our analysis, we required the transcript to be in the same orientation as its downstream gene and to be expressed under the same growth conditions. We assume that the transcript should be \geq 70 nucleotides (nt) to encode a 5' UTR 5' UTR, slightly longer than the expected 50 to 60 nts of a promoter and that the transcript extends close to the downstream genes translational start site, i.e., the transcript should extend to the penultimate or ultimate probe in the probe set of the Ig region. Figure 2 shows an example for the transcribed but not translated leader sequence of the *ompA* mRNA (Chen, L. H., Emory, S. A., Bricker, A. L., Bouvet, P. & Belasco, J. G. Structure and function of a bacterial mRNA stabilizer: analysis of the 5' untranslated region of *ompA* mRNA. *J Bacteriol* 173, 4578-86. (1991)). The PM minus MM probe intensities and the probe locations were used to determine the transcriptional start site, which was found to be close to the predicted promoter location for the *ompA* gene. A conservative set of 353 transcripts which met our expression criteria for 5' UTRs 5' UTRs were identified. 294 of these transcripts either showed concordant expression with their downstream ORF in all 13 experiments or else showed homology to *Salmonella typhi* with an E-value <0.01 (Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-402. (1997)) and an overall identity of $>65\%$. Fifteen 5' UTRs 5' UTRs contain conserved regulatory sequences, (http://www.oifn.unam.mx/Computational_Genomics/GETools/E.coli_predictions.html) (see for example, Gene Expression Analysis tools or GETtools on the website of the Nitrogen Fixation Centre of the National Autonomous University of Mexico), two match previously identified small RNAs (*sraB*, *crpT*) (Rivas, E., Klein, R. J., Jones, T. A. & Eddy, S. R. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 11, 1369-73. (2001); Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 15, 1637-51. (2001); Argaman, L. *et al.* Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 11, 941-50. (2001)) and 49 transcripts fall into potential small ORF regions.

Please replace paragraph [0075] on page 26 with the following amended paragraph:

Serial No.: 10/763,614
Attorney Docket: 3394.2

The classification of transcripts as 3-prime UTRs is analogous to that of the 5' UTRs ~~5'UTRs~~. The Ig transcript is in the same orientation as its upstream gene and expressed under the same growth conditions. In addition, we restricted the transcripts to be at least 70 bp in length, and to extend close to the upstream gene's ~~gene's~~ predicted translational stop site. According to this criteria, 122 potential ~~3'UTRs~~ 3'UTRs were identified, of which 69% are either concordantly expressed with their upstream gene in all 13 experiments or have sequence homology to *Salmonella typhi* with an E-value of <0.01 and an overall identity of >65 % (Table 5, supplemental data). Eleven of the 122 transcripts fell into potential small ORF regions.

Please replace paragraph [0076] on page 27 with the following amended paragraph:

Finally, 334 transcripts longer than 70 bp were identified. The transcripts were expressed according to the criteria but that could not be classified as operon elements, ~~5'UTRs or 3'UTRs~~ 5'UTRs or 3'UTRs based on the specific criteria for this example. This group of transcripts has a hybridization signal separate from and discontinuous with the signals from neighboring ORFs. Over 200 transcripts in this group showed sequence homology with *Salmonella typhi* or considerable expression levels (more than 3 times background). This group also contains 17 known sRNA transcripts and 31 potential new ORF regions.